Computing Trust from Revision History

Honglei Zeng[1], Maher A. Alhossaini[1], Li Ding[2],
Richard Fikes[1] and Deborah L. McGuinness[1]
[1]Knowledge Systems, AI Lab, Department of Computer Science,
Stanford University, California, USA
{hlzeng,maherhs,fikes,dlm}@ksl.stanford.edu
[2]Department of Computer Science,
University of Maryland in Baltimore County, Baltimore, Maryland, USA
dingli1@umbc.edu

Abstract

A new model of distributed, collaborative information evolution is emerging. As exemplified in Wikipedia, online collaborative information repositories are being generated, updated, and maintained by a large and diverse community of users. Issues concerning trust arise when content is generated and updated by diverse populations. Since these information repositories are constantly under revision, trust determination is not simply a static process. In this paper, we explore ways of utilizing the revision history of an article to assess the trustworthiness of the article. We then present an experiment where we used this revision history-based trust model to assess the trustworthiness of a chain of successive versions of articles in Wikipedia and evaluated the assessments produced by the model.

Keywords: Mining Revision History, Trust Computation, Collaborative Information Systems, Wikipedia

1 Introduction

If users are going to rely on information they receive from a third party, they need to have reasons to believe that the information is trustworthy. In collaborative information repositories such as Wikipedia [1], even if a document was considered trustworthy in the past, it may not still be trustworthy if it has been changed. Since these information repositories are constantly under revision, trust determination is not simply a static process. Fortunately, collaborative information repositories often maintain complete revision histories (change logs) of all documents. In the work reported in this paper, we have explored the hypothesis that revision information can be used to compute a measure of trustworthiness of revised documents. Based on that hypothesis, we developed a revision history-based trust model for computing and tracking the trustworthiness of the documents in collaborative information repositories. We represent our trust model in a dynamic Bayesian network (DBN) because a DBN is a powerful framework for modeling processes that evolve dynamically over time, which in our case, are ever changing articles.

Wikipedia is a free web-based encyclopedia and is an interesting example of collaborative information repositories where many people work in a distributed manner to create and maintain a repository of shared content. In this paper, we ground our work in Wikipedia because the trustworthiness of the articles in Wikipedia is an important and practical issue as Wikipedia continues to grow in popularity and use. Additionally, Wikipedia provides rich and accessible revision information to evaluate our approach. Wikipedia users made approximately 41 million revisions, an average of 12 versions per article, from July 2002 to January 2006.

The rest of our paper is structured as follows. We introduce the concept and intuitions of revision trust in section 2. We develop a dynamic Bayesian network for our revision history-based trust model in section 3. In section 4, we describe experimental results used to evaluate the method and discuss their implications. We discuss related work in section 5 and conclude our paper with a discussion of future work in section 6.

2 Trust Issues in Wikipedia

Wikipedia is a popular online encyclopedia which is collaboratively written and maintained by volunteers world wide. As of January 2006, it has more than 3.3 million articles in 200 languages (970,000 articles in Eng-

[1] www.wikipedia.com

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE
Computing Trust from Revision History

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Maryland in Baltimore County,Department of Computer Science,Baltimore,MD,21250

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **8**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

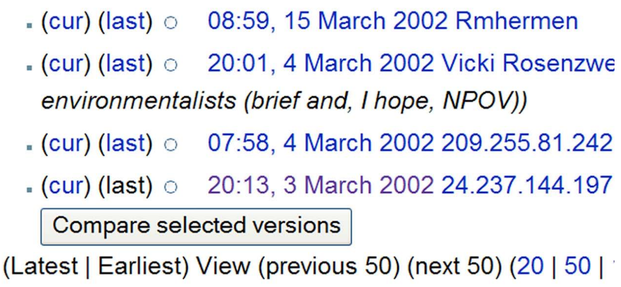
lish) [2]. Wikipedia is now among the top 30 most visited websites according to the web traffic statistics provided by Alexa.com.

As Wikipedia continues to expand in content and use, issues concerning the trustworthiness of information grow. While recent studies (e.g. [1]) show that the science articles in Wikipedia are generally trustworthy, there have been several discoveries of inaccurate or biased articles. Although many approaches have been tried to address the trust issue in Wikipedia, destructive user actions cannot be completely prevented due to Wikipedia's open editing policies that allow anyone to freely create and edit articles.

While manual monitoring on Wikipedia has worked fairly well, we consider it critical to build computational trust models for Wikipedia and leverage automated trust management to complement manual monitoring. It is preferable but manually infeasible to annotate trustworthiness at the article level, because Wikipedia as a whole cannot be completely trusted. In addition, even if an article was considered trustworthy in the past, it may not still be trustworthy if it has been changed. Automated trust models are needed to monitor changes in trustworthiness of articles caused by revisions.

2.1 Revision history-based trust model

This paper investigates automated trust computation using the revision history of an article. Throughout this paper, a *revision* on an article refers to the action of editing the article by an author. When an article is revised, a new version of the article is created to archive the revised content. Thus, the i^{th} version of an article is the article after i revisions[3]. A *revision history* of an article is a sequence of its versions ordered by their creation time. Wikipedia archives complete revision history of its articles; for example, Figure 1 shows the first four versions of the article *U.S. National Forest*.



- (cur) (last) ○ 08:59, 15 March 2002 Rmhermen
- (cur) (last) ○ 20:01, 4 March 2002 Vicki Rosenzwe
  environmentalists (brief and, I hope, NPOV))
- (cur) (last) ○ 07:58, 4 March 2002 209.255.81.242
- (cur) (last) ○ 20:13, 3 March 2002 24.237.144.197

(Latest | Earliest) View (previous 50) (next 50) (20 | 50 |

Figure 1. A snapshot of the revision history of the article *U.S. National Forest* **in Wikipedia**

A revision history-based trust model, or simply revision trust model, is built on our hypothesis that revision information can be used to compute a measure of trustworthiness: trustworthiness of the revised version depends on the trustworthiness of the previous version, the author of the last revision, and the amount of text involved in the last revision.

A revision trust model may help address a list of interesting trust problems including the following:

- Article trust: trustworthiness of a version of an article;
- Fragment trust: trustworthiness of a fragment in a version of an article (an article may consist of fragments contributed by different authors);
- Author trust: trustworthiness of an author; in particular, author trust for specific domains. For example, it is possible to derive an author's trustworthiness with respect to the domain of "Wine" with over 500 revisions of the Wikipedia article *Wine* or the trustworthiness of an author with respect to the domain of "Brandy" with 150 revisions of the Wikipedia article *Brandy*. Such rich revision information may provide enough data for algorithms to infer domain-specific trust to a finer granularity than with other approaches.

As a first step towards a comprehensive revision trust model, we focus on article trust in this paper, though other trust issues, i.e. fragment trust and author trust, are also under our investigation.

3 A Revision Trust Model for article trust

We developed a revision trust model to compute article trust and represented the model using a dynamic Bayesian network (DBN).

There have been extensive studies of the approaches to quantifying trustworthiness (of entities). For example, Golbeck and Hendler [2] defined a binary scale for trust values (trusted or not trusted). In this paper, trust values are defined over a continuous range $[0, 1]$, as has been done in multiple previous approaches (e.g., [5] and [4]). We use 0 to represent complete untrustworthiness and 1 to represent complete trustworthiness. The trust value of an article is interpreted as the percentage of the content in the article that is trustworthy; for example, a trust value of 0.6 means 60% content of the article is trustworthy. Similarly, an author with a trust value 0.6 means 60% of the content he or she writes is trustworthy.

This trust representation might be too simple to fully capture the complexity of trustworthiness in Wikipedia. Nevertheless, it has low computational overhead and is intuitive to end users of Wikipedia (when the results are presented visually to them). Our revision trust model may

[2] en.wikipedia.org/wiki/Wikipedia

[3] We define the original article to be the 0^{th} version.

be extended to work with other, potentially more complex, trust models.

3.1 Notation

We view a revision as a collection of *deletions*, each of which removes some content from an article, and *insertions*, each of which adds some content to an article[4]. Given an article with n revisions (thus with $n+1$ versions), we use the following notation throughout this paper:

V_i refers to the i^{th} version of an article which resulted from a revision applied to the previous version V_{i-1}. V_0 is the original article and V_n is the final version. t_{V_i} is the trust value of V_i.

A_i refers to the author who revised V_{i-1}. A_0 is the creator of the original article. t_{A_i} is the trust value of A_i.

I_i and D_i refer to the inserted content and deleted content from and to V_i by A_{i+1} respectively. If x is a piece of text, then $|x|$ denotes the size (i.e. number of words) of x. It is easy to see that $0 \leq |I_i| \leq |V_{i+1}|$ and $0 \leq |D_i| \leq |V_i|$.

3.2 A Dynamic Bayesian Network Trust Model

We use Dynamic Bayesian networks (DBN) to model the evolution of article trust over revisions. The DBN is defined by a pair (B_s, B_o) where B_s is the graph structure of the network and B_o is the set of the network's conditional density distributions.

3.2.1 Graph structure B_s

Figure 2 shows a segment of B_s from version V_i to V_{i+1}. The segment is repeated n times in B_s (i from 0 to $n-1$) to model a sequence of $n+1$ versions of an article.

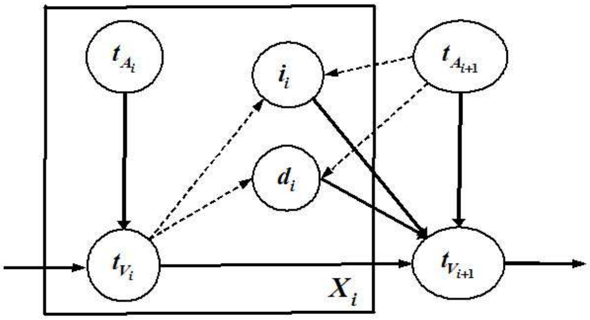


Figure 2. DBN segment from V_i to V_{i+1}.

The state of the DBN $\mathbf{X_i}$ at the i^{th} revision is represented as a quad $(t_{V_i}, t_{A_i}, i_i, d_i)$, where i_i is the amount of the insertion, and d_i is the amount of the deletion. t_{V_i} and t_{A_i} are continuous variables over $[0,1]$, while i_i and d_i are continuous variables over $[0,\infty]$. Our DBN satisfies the Markov property: $f(\mathbf{X_{i+1}}|\mathbf{X_i}, \mathbf{X_{i-1}}, ..., \mathbf{X_0}) = f(\mathbf{X_{i+1}}|\mathbf{X_i})$.

Additionally, the solid arrows in Figure 2 encode the dependency relationships in the DBN segment. The trustworthiness of an article version is dependent on the trustworthiness of the previous version and the author of the last revision, the amount of the insertion, and the amount of the deletion.

There may be dependencies between i_i and t_{V_i} and $t_{A_{i+1}}$, and between d_i and t_{V_i} and $t_{A_{i+1}}$, as indicated by the dotted arrows in Figure 2. For example, a trustworthy author is likely to make a large amount of changes to a very untrustworthy article. This paper assumes independencies between i_i, d_i and t_{V_i}, $t_{A_{i+1}}$. Our assumption might be an issue to debate; nevertheless, i_i and d_i are also dependent on other important factors, such as the interest an author has in an article. Since these factors are not modeled in the DBN, deciding the conditional density functions between i_i, d_i and t_{V_i}, $t_{A_{i+1}}$ can be difficult and less useful. Although we are still investigating approaches to remove the independence assumption, good results have been observed in our experiments even with this assumption.

We seek to determine the posterior density distribution of $f(t_{V_{i+1}})$. Given the dependencies in the graph structure B_s, B_o is fully characterized by $f(t_{V_0}|t_{A_0})$ and $f(t_{V_{i+1}}|t_{V_i}, t_{A_{i+1}}, i_i, d_i)$.

3.2.2 $f(t_{V_0}|t_{A_0})$

The trustworthiness of the original article is only dependent on its author in our model. If we assume the dependency between t_{V_0} and t_{A_0} is deterministic, the conditional probability is: $P(t_{V_0} = a_0 | t_{A_0} = a_0) = 1$.

That is, if an author is 0.8 trustworthy, then an article written by that author is 0.8 trustworthy with probability 1. Nevertheless, in reality, that article is very unlikely to be precisely 0.8 trustworthy. For example, the trustworthiness of the article mentioned above can be in a range of $0.75 - 0.85$ with high probability. The uncertainty in the trustworthiness is caused by uncertain factors such as the context where an author writes an article.

A common approach is to assume normality and model conditional density functions with Gaussian distributions. In this paper, we chose beta distributions because trust variables are defined in the range $[0,1]$ where beta distributions are also defined. Nevertheless, we do not view revisions as independent events, even though beta distribution is normally associated with binomial process.

3.2.3 Beta distributions

The beta distributions are a family of distributions with two parameters α and β.

[4] In this paper, we view an update as equivalent to a deletion followed by an insertion, ignoring their subtle differences.

$beta(p|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$, where $B(x,y)$ is the beta function: $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$.

$beta(p|\alpha,\beta)$ can take on different shapes depending on the values of α and β, as depicted in Figure 3. When $\alpha, \beta > 1$, the curve is a desired unimodal for modeling the uncertainty in the trust distribution. We chose $\alpha, \beta \geq 10$ in this work as we make a simplifying assumption that the variance σ of the distribution should be neither too small nor too large when the mean μ is close to 0.5. On the other hand, if μ is close to 0 or 1, σ does not make much difference in the choices of α and β because σ is bounded by a very small value $\mu(1-\mu)$. Given this constraint and the fact that the mean of $beta(p|\alpha,\beta)$ is $\mu = \frac{\alpha}{\alpha+\beta}$, we have

$$\begin{aligned} &\alpha = 10, \beta = \frac{10-10\mu}{\mu}; or \\ &\alpha = \frac{10\mu}{1-\mu}, \beta = 10. \end{aligned} \tag{1}$$

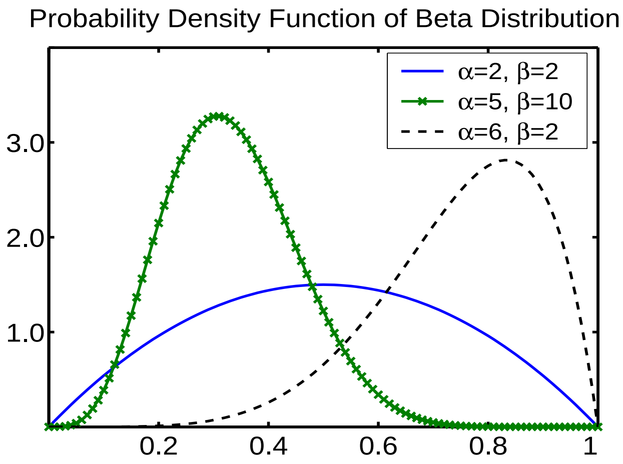


Figure 3. Shapes of beta distributions with different α and β. The mean of $beta(p|\alpha,\beta)$ is $\mu = \frac{\alpha}{\alpha+\beta}$ and the variance is $\sigma = \frac{\alpha\beta}{(\alpha+\beta)^2((\alpha+\beta+1)}$.

The distribution of t_{V_0} given t_{A_0} is therefore

$$f(t_{V_0}|t_{A_0} = a_0) = beta(p|\alpha_0, \beta_0) \tag{2}$$

We choose $\mu_0 = a_0$, then α_0 and β_0 can be determined by Equation 1.

3.2.4 $f(t_{V_{i+1}}|t_{V_i}, t_{A_{i+1}}, i_i, d_i)$

We now consider the conditional density distribution of $t_{V_{i+1}}$ in the DBN, for $0 \leq i \leq n-1$; similarly to $f(t_{V_0}|t_{A_0})$, we assume that these distributions are also beta distributions:

$$\begin{aligned} &f(t_{V_{i+1}}|t_{V_i} = t, t_{A_{i+1}} = a_{i+1}, i_i = |I_i|, d_i = |D_i|) \\ &\quad = beta(p|\alpha_{i+1}, \beta_{i+1}) \end{aligned} \tag{3}$$

Given Equation 1, we only need to compute μ_{i+1} to fully determine $beta(p|\alpha_{i+1}, \beta_{i+1})$.

In our approach, the trustworthy portion of V_{i+1} is the trustworthy portion of V_i plus the trustworthy insertion of A_{i+1} minus the trustworthy portion that A_{i+1} incorrectly removed from V_i.

We know that $t|V_i|$ of V_i is trustworthy and $(1-t)|V_i|$ is untrustworthy. Additionally, since $t_{A_{i+1}} = a_{i+1}$, we expect $a_{i+1}|I_i|$ of I_i to be trustworthy and $(1-a_{i+1})|I_i|$ to be untrustworthy.

The consequences of the deletions are complicated. In this paper, we consider the following four scenarios about deletions[5]:

- $(1-a_{i+1})|D_i| \leq t|V_i|$ and $a_{i+1}|D_i| \leq (1-t)|V_i|$. A_{i+1} expects to (incorrectly) remove $(1-a_{i+1})|D_i|$ of the trustworthy portion of V_i. In this case, since this amount removed is no more than the amount of the existing trustworthy portion in V_i (i.e. $t|V_i|$), A_{i+1} actually removes $(1-a_{i+1})|D_i|$ of the trustworthy portion of V_i;

- $(1-a_{i+1})|D_i| > t|V_i|$ and $a_{i+1}|D_i| \leq (1-t)|V_i|$. A_{i+1} (incorrectly) removes all the trustworthy portion of V_i;

- $(1-a_{i+1})|D_i| \leq t|V_i|$ and $a_{i+1}|D_i| > (1-t)|V_i|$. A_{i+1} (correctly) removes all the untrustworthy portion of V_i (i.e. $(1-t)|V_i|$); thus, A_{i+1} has to remove an additional $a_{i+1}|D_i| - (1-t)|V_i|$ amount of the trustworthy portion of V_i. The total amount of the trustworthy portion that A_{i+1} removes from V_i is thereby $(1-a_{i+1})|D_i| + (a_{i+1}|D_i| - (1-t)|V_i|)$.

- $(1-a_{i+1})|D_i| > t|V_i|$ and $a_{i+1}|D_i| > (1-t)|V_i|$. This case is impossible because $|D_i| \leq |V_i|$.

In summary, μ_{i+1}, the mean of $beta(p|\alpha_{i+1}, \beta_{i+1})$, is the size of the trustworthy portion in V_{i+1} divided by the total size of V_{i+1}, where $|V_{i+1}| = |V_i| + |I_i| - |D_i|$. We combine the above four cases into Equation 4.

$$\begin{aligned} \mu_{i+1} = \{&t|V_i| + a_{i+1}|I_i| - \min((1-a_{i+1})|D_i|, t|V_i|) \\ &- \max(a_{i+1}|D_i| - (1-t)|V_i|, 0)\}/|V_{i+1}| \end{aligned} \tag{4}$$

Given μ_{i+1} in Equation 4, α_{i+1} and β_{i+1} are determined by Equation 1, and thus the conditional distribution in Equation 3 is determined. Our DBN is now complete with both B_s and B_o fully defined.

[5] Our method is not the only approach to determine the trustworthy portion that A_{i+1} incorrectly removed from V_i.

4 Experiments

We now describe the experiments and their results. The inference computations in our DBN were carried out using the freely available software BUGS (Bayesian inference Using Gibbs Sampling [11]). BUGS is a software package for BN inference using the Markov Chain Monte Carlo (MCMC) method.

4.1 Experiment Settings

In order to evaluate our trust model for computing article trust, we collected a set of English articles from the Geography category in Wikipedia in January 2006. We chose articles from the same category so that their trustworthiness values could be viewed as comparable.

Evaluation of trust algorithms in Wikipedia is difficult because the trustworthiness of Wikipedia articles is not explicitly asserted and is subject to personal opinions. To obtain the reference trust value for each article, we consider three manually classified groups of articles in Wikipedia:

- *featured articles*[6] which are considered highly trustworthy in the Wikipedia community because they have been thoroughly reviewed for style, prose, completeness, accuracy and neutrality;
- *clean-up articles*[7] which are considered untrustworthy because they have been marked for major revision by Wikipedia authors;
- *normal articles* which are the remaining articles.

Our data set consisted of 50 featured articles, 50 clean-up articles, and 768 normal articles (a total of 40450 revisions). The number of featured articles and clean-up articles we considered in the experiments is relative small because only 0.1% of Wikipedia articles are featured articles and 1.3% are clean-up articles. In particular, there are less than 80 featured articles in the Geography category.

Author trust is synthesized based on the background knowledge from Wikipedia. Although there are no explicit trust values associated with the authors in Wikipedia, we can provide coarse approximations based on the editing privileges of the authors. Currently, Wikipedia supports four levels of authorship: administrators (including stewards and developers), registered authors, anonymous authors, and blocked authors, with each level having decreasing editing powers and trustworthiness. We thereby use the following beta distributions to approximate the trustworthiness of the authors : $beta(p|190, 10)$ for administrators, $beta(p|23, 10)$ for registered authors, $beta(p|15, 10)$ for anonymous authors, and $beta(p|10, 190)$ for blocked authors. The means of the distributions are 0.95, 0.7, 0.6, and 0.05 respectively. The exact values are chosen rather arbitrarily; however, the relative differentials between the trust values reflect our assessment of the trustworthiness of the authors with different editing privileges.

The amount of insertion and deletion of each revision can be calculated by comparing consecutive article versions using a *diff* algorithm. Our implementation of diff was based on the well known *longest common subsequence* algorithm [9]; and the differences between versions were counted at word level.

Table 1. The statistics of data set

	featured articles	clean-up articles	normal articles
Average number of revisions	726	56	27
Average percentage of administrators	29.4%	12.5%	28.4%
Average percentage of registered authors	45.2%	56.6%	52.1%
Average percentage of anonymous authors	24.1%	29.6%	18.7%
Average percentage of blocked authors	1.3%	1.3%	0.8%
Average size of the final version	3385	524	274
Average percentage of insertion per revision	2.5%	6.3%	6.2%
Average percentage of deletion per revision	2.1%	4.7%	2.4%

Table 1 shows the statistics of the data set. Featured articles have far more revisions than clean-up articles and normal articles, while clean-up articles have the lowest percentage of administrator authors. Though the number of revisions and the trustworthiness of authors are the two most important factors in determining the trustworthiness of an article, other factors, such as the amount of the changes and the order of the revisions are also important. It is interesting to note that normal articles have the lowest number of revisions, which shows that both featured articles and clean-up articles receive more attentions from Wikipedia authors.

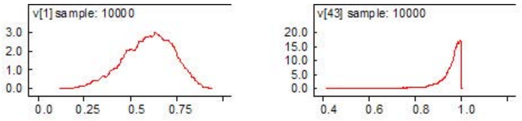
4.2 Evaluation and Discussion

We defined and ran our trust model in BUGS and obtained posterior density distributions $f(t_{V_i})$ for article versions in our data set. For example, Figure 4 shows the trust density distributions of V_0 and V_{43} of the article *U.S.*

[6] en.wikipedia.org/wiki/Wikipedia:Featured_articles

[7] en.wikipedia.org/wiki/Clean-up

National Forest [8]. *U.S. National Forest* was created by an anonymous author, thus the distribution of t_{V_0} is similar to the trust distribution of anonymous authors. The final version was highly trustworthy after 43 revisions. In most cases, the trustworthiness measure of article versions is increasingly precise with more revisions; for example, as shown in Figure 4, the variance of $f(t_{V_{43}})$ is much smaller than that of $f(t_{V_0})$;



(a) Distribution of t_{V_0} **(b) Distribution of $t_{V_{43}}$**

Figure 4. Trust density distributions of the first version and the final version of the article *U.S. National Forest***.**

We are most interested in the mean of $f(t_{V_n})$, which we refer as $\overline{\mu_n}$ [9]. $\overline{\mu_n}$ indicates the trustworthiness of the latest version of an article. The results are summarized in Table 2. Featured articles have the highest average of $\overline{\mu_n}$ and the lowest variance, while clean-up articles are the opposite. Note that even though several simplifying assumptions were made in this work, our model showed significant differences of trustworthiness between featured articles and clean-up articles.

Table 2. The average and variance of $\overline{\mu_n}$ of 50 featured articles, 50 clean-up articles and 768 normal articles.

	featured articles	clean-up articles	normal articles
Average of $\overline{\mu_n}$	0.885	0.768	0.808
Variance of $\overline{\mu_n}$	0.011	0.019	0.016

The values in Table 2 are the trust values of the final versions (i.e. the most recent versions) of featured articles and clean-up articles. The average $\overline{\mu}$ when featured articles were initially approved is 0.903, just slightly higher than the current average 0.885. On the contrary, the average $\overline{\mu}$ when clean-up articles were marked is 0.739, lower than the current value 0.768. We found that six clean-up articles (out of 50) had been dramatically improved since they were marked as clean-up articles. For example, $\overline{\mu}$ of article *Victoria Park (Hartlepool)* jumped from 0.705 to 0.884 within a month. It seems that typically there are small changes of trustworthiness in featured articles after they became featured articles, while clean-up articles are being continuously improved.

We developed a classifier based on the aforementioned 50 featured articles and 50 clean-up articles in Table 2 to separate featured articles and clean-up articles. The training set contains 100 pairs (x, y), where x is the trust value of an article and y is its class. In this data set, we only consider the revision history of a featured article (or a clean-up article) up to the point where it was initially approved by Wikipedia authors. We do not use the complete revision history which includes the most recent article versions whose trustworthiness may be unknown. Since we use the trust value to predict a class that an article belongs to, the learned rule for featured article is: $x > 0.842$. Thus in practice we consider an article trustworthy if its trust value is higher than 0.842. A test size of 200 new articles (48805 article revisions) was evaluated. The percentage for correctly predicting featured articles is 82% and that for clean-up articles is 84%.

We studied citation-based trust in [8]. Citation-based trust algorithms are a family of algorithms that derive trust based on the citation relationships among articles. For example, a well-cited article may be more trustworthy than an article that has no citation. We showed that citation-based trust algorithms may not be very effective for computing trustworthiness of assertions in an aggregated knowledge repository such as Wikipedia. In particular, the percentage for correctly predicting featured articles is 46% and that for clean-up articles is 54% based on a classifier developed from a citation-based trust algorithm. Clearly, revision history-based trust algorithm significantly improves the accuracy of trust computation.

Based on manual inspection, our model appears correct in estimating the changes of trustworthiness caused by major revision events (e.g. a revision with more than 10% insertion and/or deletion). For example, in Figure 5, we examine every $\overline{\mu_i}$ from the first version to the final version of *U.S. National Forest*. We manually identified six major revisions from the revision history of the article and depicted these events in Figure 5. Our trust model successfully estimated the consequences of five events (which are depicted in white boxes). The only exception is event four, where our model predicted a decrease of trustworthiness due to a 15% insertion made by a blocked author. Nevertheless, our manual analysis showed that this author's insertion here was genuine. Our results may be improved by developing author trust models for modeling complicated author trust behaviors. For example, lack of expertise (blocked status) in one area does not mean a lack of expertise (blocked status) in another.

Additionally, we have the following observations: (1)

[8] en.wikipedia.org/w/index.php?title=U.S._National_Forest

[9] $\overline{\mu_n}$ is the mean of the posterior density distribution $f(t_{V_n})$, while in Equation 4, μ_n is the mean of the conditional distribution $f(t_{V_n}|t_{V_{n-1}}, t_{A_n}, i_{n-1}, d_{n-1})$.

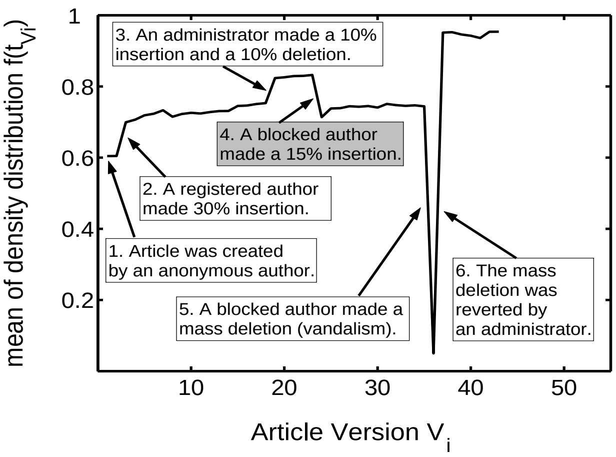


Figure 5. Changes in trustworthiness of all versions of *U.S. National Forest*. **These events except event four are consistent with the results of our model.**

The changes of trustworthiness are rather smooth unless a major revision occurs; (2) In event 5, a large drop in trustworthiness was caused by a mass deletion, which was subsequently reverted by an administer in event 6. The trustworthiness of the reverted version might be considered to be the same as the version before the vandalism occurred, but since the revision was performed by an administrator, the trustworthiness was higher in our model.

5 Related Work

Lih [6] studied the correlation between the numbers of revisions and unique authors to Wikipedia articles and the quality of these articles. Voss [14] conducted a comprehensive analysis on various aspects of Wikipedia articles. Viégas et al. [13] presented a tool that visualized revision history flow and through which they revealed some interesting revision patterns in Wikipedia. While qualitative studies are important, we believe our approach is the first computational trust model utilizing revision history data in Wikipedia.

Trust in P2P and social networks has been extensively studied in recent years [5], [4] and [3]. Unlike the typical social networks, the trustworthiness of authors and articles are not explicitly asserted in Wikipedia. Revision trust is substantially different from those approaches which are based on the transitivity property of trust.

While our current use of Bayesian networks is modest, the reader can refer to [10] for more background on Bayesian networks and DBNs and envision future extensions to our work, e.g., learning α_i and β_i parameters from revision history.

Other related work includes mining web logs (e.g., [12]) and mining software revision history (e.g., [7]). Most such applications were built on mining association rules. However, association rules or simple revision parameters (such the number of revisions) are not very useful in computing and tracking trustworthiness of articles that are under constant changes. For example, a featured article could become untrustworthy if it has been changed despite the fact that the number of revisions is monotonically increasing.

6 Conclusions

Trust in collaborative information repositories is becoming increasingly important as people are relying more on online information and are more actively participating in online collaborations. In this paper, we made the following contributions towards the understanding and computing of trust in collaborative information repositories, in particular, Wikipedia. We introduced the concept of revision history-based trust and developed a dynamic Bayesian network trust model that utilized rich revision information in Wikipedia. Our experiments showed promising results, even though we made several simplifying assumptions in this work. We showed an evaluation method in Wikipedia based on its feature articles, clean-up articles and the levels of author editing privileges. Our work provided a methodology for comparing and evaluating future computational trust algorithms.

Based on our DBN model, we believe the reasons for Wikipedia being generally trustworthy are: (1) most Wikipedia authors seem to have good intentions (there are only 1.3% blocked authors); (2) Wikipedia administrators have the responsibility and authority to settle disputes, prevent vandalism, and block inappropriate authors. While there are a small number of administrators (0.09%), they have made much larger contributions to Wikipedia, for example, 29.4% revisions of featured articles were made by administrators in our experiments, according to Table 1; (3) Wikipedia maintains a complete revision history of articles from which a previous content modification can be easily reverted.

The benefits of revision trust to Wikipedia users are significant. Visualization of the computed trust values may help users to decide what information they should trust. Users may also have the option to view the most trustworthy version of an article, in addition to the most recent one. Furthermore, revision trust can improve Wikipedia's quality control process; for example, our model provides an appealing approach to monitoring changes in trustworthiness and thereby providing timely notifications of vandalism and other forms of malicious content modifications.

Our work may be extended in several directions: (1) The article trust model may be refined and improved; we expect better performance when the simplifying assumptions in our model are removed; (2) Revision trust may help to solve many difficult trust issues, e.g., fragment trust and author trust; (3) Though we focused on computing trust in Wikipedia, it is interesting to extend revision trust to other collaborative systems with rich revision information.

7 Acknowledgments

This research was largely supported by Stanford's DARPA contract #HR0011-05-1-0019-P00001 and DTO contract #2003*H278-000*000. We would also like to thank Paulo Pinheiro da Silva, Mayukh Bhaowal and Dhyanesh Narayanan for valuable conversations and help in implementation.

References

[1] J. Giles. Internet encyclopaedias go head to head. In *Nature 438, 900-901 (15 Dec 2005)*, 2005.

[2] J. Golbeck and J. Hendler. Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *In Proceedings of EKAW'04*, 2004.

[3] J. Golbeck and J. Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer Communications and Networking Conference*, 2006.

[4] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM Press, 2004.

[5] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, 2003.

[6] A. Lih. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th International Symposium on Online Journalism*, 2004.

[7] V. B. Livshits and T. Zimmermann. Dynamine: finding common error patterns by mining software revision histories. In *ESEC/SIGSOFT FSE*, pages 296–305, 2005.

[8] D. L. McGuinness, H. Zeng, P. Pinheiro da Silva, L. Ding, D. Narayanan, and M. Bhaowal. Investigations into trust for collaborative information repositories. In *The Workshop on the Models of Trust for the Web (MTW'06)*, 2006.

[9] E. W. Myers. An o(ND) difference algorithm and its variations. *Algorithmica*, 1(2):251–266, 1986.

[10] R. E. Neapolitan. In *Learning Bayesian Networks*. Prentice Hall, 2004.

[11] D. J. Speigelhalter, A. Thomas, N. G. Best, and W. R. Gilk. Bugs: Bayesian inference using gibbs sampling, version 0.50. In *http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml*, 2005.

[12] R. Srikant and Y. Yang. Mining web logs to improve website organization. In *the Ninth International World Wide Web Conference*, pages 430–437, 2001.

[13] F. Viéas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of CHI04*, 2004.

[14] J. Voss. Measuring wikipedia. In *In Proceedings 10th International Conference of the International Society for Scientometrics and Informetrics*, 2005.